

Extended Generalised Pareto Models for Tail Estimation

Ioannis Papastathopoulos* and Jonathan A. Tawn

November 30, 2011

Abstract

The most popular approach in extreme value statistics is the modelling of threshold exceedances using the asymptotically motivated generalised Pareto distribution. This approach involves the selection of a high threshold above which the model fits the data well. Sometimes, few observations of a measurement process might be recorded in applications and so selecting a high quantile of the sample as the threshold leads to almost no exceedances. In this paper we propose extensions of the generalised Pareto distribution that incorporate an additional shape parameter while keeping the tail behaviour unaffected. The inclusion of this parameter offers additional structure for the main body of the distribution, improves the stability of the modified scale, tail index and return level estimates to threshold choice and allows a lower threshold to be selected. We illustrate the benefits of the proposed models with a simulation study and two case studies.

Keywords: extreme value theory; extended generalised Pareto distribution; tail estimation; threshold selection; liver toxicity

1 Introduction

The area of extreme value theory focuses on the study and development of stochastic models that can be used for inference on applied problems related to the frequency of very big (or very small) values in random experiments. One such widely used model is the generalised Pareto (GP) distribution defined by its distribution function

$$F(x; \boldsymbol{\lambda}) = 1 - (1 + \xi x / \sigma)_+^{-1/\xi}, \quad x > 0, \quad (1)$$

where $\boldsymbol{\lambda} = (\sigma, \xi)$ is a vector of parameters in $(0, \infty) \times (-\infty, \infty)$ and $z_+ = \max(z, 0)$. Consider a random variable X arising from an absolutely continuous distribution function F_X and let also $x^{F_X} = \sup\{x : F_X(x) < 1\}$ be the upper end point of F_X . Pickands (1975) shows that if there exists a scaling function $h_X(u) : \mathbb{R} \rightarrow \mathbb{R}_+$, $u < x^{F_X}$, such that the scaled excess random variable $\{(X - u)/h_X(u)\} | X > u$ converges in distribution to a non-degenerate limit as $u \rightarrow x^{F_X}$, then this is necessarily of the same type as the GP distribution, i.e.,

$$\lim_{u \rightarrow x^{F_X}} \Pr \left\{ \frac{X - u}{h_X(u)} < x \mid X > u \right\} = F(x; \boldsymbol{\lambda}_0), \quad x > 0, \quad \boldsymbol{\lambda}_0 = (1, \xi). \quad (2)$$

Without loss of generality, the scaling function h_X can be defined by the reciprocal hazard function of X , i.e., $h_X(u) = \{1 - F_X(u)\}/F'_X(u)$. Pickands (1986) shows that a necessary and sufficient condition for twice differentiable convergence is $\lim_{u \rightarrow x^{F_X}} h'_X(u) = \xi$, meaning that

*Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK. Email: i.papastathopoulos@lancaster.ac.uk and j.tawn@lancaster.ac.uk

not only limit (2) holds but the corresponding densities and derivatives of densities converge. The parameter ξ is most commonly referred to as the shape parameter or the tail index of the distribution and we adopt the latter since we use the word shape for a different characteristic in the paper. The sign of the tail index ξ indicates the decay of the tail of F_X : $\xi > 0$ means that F_X has a heavy-tailed distribution, $\xi \rightarrow 0$ corresponds to exponential decay and $\xi < 0$ means that F_X has finite upper end point, i.e., $x^{F_X} < \infty$.

Consider a sequence of independent and identically distributed measurements $\{x_1, \dots, x_n\}$ arising from the random variable X . Standard practice in applications where we want to estimate extreme quantiles of the underlying distribution is to follow the approach of Davison and Smith (1990) and assume that the limit relationship (2) holds exactly for some threshold $u < \max_{i=1, \dots, n}(x_i)$; i.e., $X - u | X > u$ has distribution function $F(x; \sigma_u, \xi)$ where the function $h_X(u)$ is absorbed in the distribution function F as a threshold dependent scale parameter $\sigma_u > 0$. Extreme events are defined by the threshold exceedances $\{x_i : x_i > u\}$ and subsequently, the GP distribution is fitted to the random sample of excesses $\{x_i - u : x_i > u\}$ using maximum likelihood techniques. The fitted model is then extrapolated to levels above which no data are observed. Two central assumptions are imposed when using this procedure in practice. One is that the asymptotic argument of equation (2) is valid for the distribution function of the data under study and second is that an appropriate threshold u can be found such that the GP model provides a good approximation to exceedances of u . For a number of reasons such as the cost and time of collecting data, few observations of a measurement process might be recorded in applications. Since $u < \max_{i=1, \dots, n}(x_i)$ and n is small, $F_X(u) \ll 1$ and so limit (2) is likely to be a poor approximation to the distribution of exceedances of u in such cases.

Figure 1 shows such an example of residual bilirubin data collected from a clinical study of 606 patients who were randomized to 4 doses of a drug, the highest dose of which is considered to have potential for liver toxicity. Data were available prior to treatment (baseline visit) and after 6 weeks of treatment (postbaseline). The residual bilirubin observations are the residuals of linear median regression models of postbaseline on the baseline in each dose, see also Southworth and Heffernan (2010) for a similar analysis.

The 4 different doses are coded here with increasing dose as A , B , C and D . The published literature reports jaundice, hepatitis and similar symptoms in approximately 1 out of 500 patients taking the dose D of this drug, see Southworth and Heffernan (2010). According to FDA (2008) joint occurrence of extremes of the total bilirubin and aminotransferase laboratory variables are indicative of drug induced liver toxicity. Therefore, proper statistical modelling of their extremes is vital for assessing the liver toxicity of a new drug. However, the limited amount of information in each dose illustrates the problems of relying on the GP distribution. We will return to the analysis of these data in § 4.2.

The issue of specifying an appropriate threshold for fitting the GP distribution constitutes the major problem of the Davison and Smith (1990) approach. Typically, a threshold u is chosen as the lowest possible value above which the estimates of the tail index ξ and the modified scale $\sigma^* = \sigma_u - \xi u$ stabilise (see Coles, 2001). Departures from the GP distribution again imply that such a threshold might not be observable. Moreover, the higher the threshold the larger the sampling variability of the estimates of the modified scale and tail index parameters which leads to estimates being unstable over different thresholds. A variety of methods have been developed in the literature to address the problem of departures from the GP model assumption. Peng (1998), Feuerverger and Hall (1999) and Beirlant et al. (1999) among others, use second order

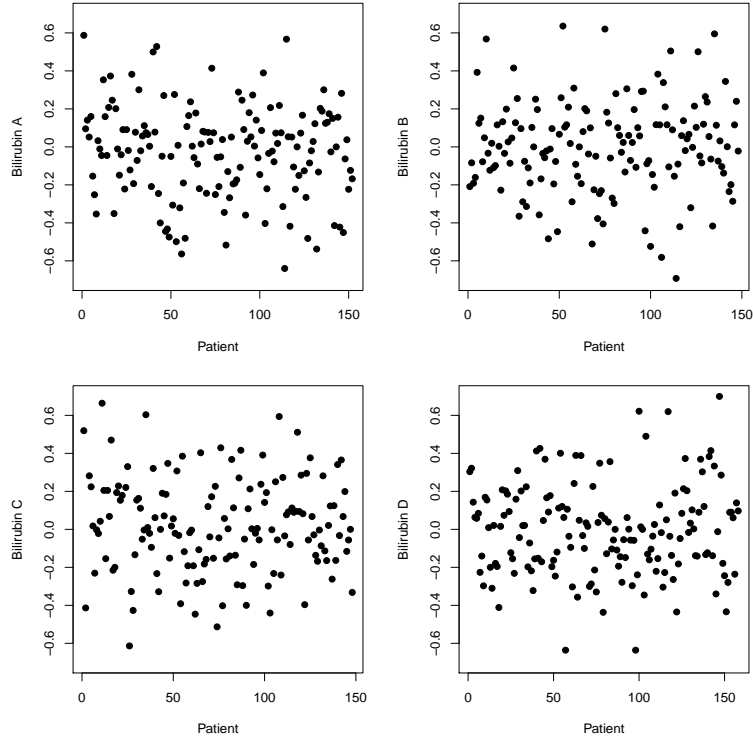


Figure 1: Residual total bilirubin variable measured from 606 patients at four different doses A , B , C and D .

refined models to select the threshold and proceed with the estimation of tail characteristics by using the GP distribution. Beirlant et al. (2009) also use a second order approach for the modelling of the tail probability as well as for the extrapolation. An even more unsettling feature arises in cases where two or more seemingly plausible thresholds yield significantly different estimates of extreme quantities of interest. In such circumstances, the threshold selection is liable to be the subjective choice of the practitioner. Frigessi et al. (2002) propose unsupervised tail estimation with the use of a dynamic mixture model aimed for the entire distribution of the data. Tancredi et al. (2006) take into account the threshold uncertainty by allowing the threshold to be estimated as a statistical parameter in a mixture model. MacDonald et al. (2011) also use a mixture model where the non-extreme part of the data is estimated non-parametrically. Wadsworth and Tawn (2011) exploit penultimate theory to model threshold uncertainty and provide a likelihood ratio testing procedure for the threshold selection.

All of the aforementioned approaches are either based on second order asymptotic arguments adding a little extra flexibility to the fit of the GP distribution or they model the entire distribution, i.e., model the body as well as the tail. Furthermore, some of these approaches are confined to the heavy-tailed case $\xi > 0$ (Beirlant et al., 2009). Our goal in this article is to construct parametric models for exceedances over thresholds that are more flexible than the GP distribution and add further insight on the threshold selection problem. For this purpose, in § 2.3 we construct a new class of probability models with distribution function $G(x; \kappa, \boldsymbol{\lambda})$, where $\kappa > 0$ is a shape parameter, that generalises the GP distribution in the sense that there exists a $\kappa^* > 0$ such that $G(x; \kappa^*, \boldsymbol{\lambda}) = F(x; \boldsymbol{\lambda})$. This parameter κ offers additional structure by inducing skewness to the GP distribution while retaining $\xi \in \mathbb{R}$ as the tail index. Thus, a class of departures from the GP assumption of limit (2) are captured by this shape parameter κ . We will show that the inclusion of κ improves the stability of the estimates of the tail in-

dex and modified scale parameter and allows a lower threshold to be selected. Consequently, extrapolations based on different thresholds are more stable than those obtained from the GP distribution, a feature which makes the choice of threshold less important.

In § 2 we present three extensions of the GP distribution and derive a characterisation of a class of models which includes these examples. The new models are given along with the description of the methodology implemented for the analysis of threshold exceedances. A statistical test aiding threshold selection is also illustrated. The effect of the new probability models on the statistical analysis of extremes is assessed with a simulation study in § 3. Finally, in § 4 we illustrate the benefits of the new models through the analysis of extreme flow data of the River Nidd, a dataset with known difficulties in threshold selection, and the clinical trial data in Figure 1.

2 Theory and Models

2.1 Notation and motivation

Here and throughout, we denote by $\beta(\cdot; a, b)$ and $\gamma(\cdot; a)$ the regularised incomplete beta and regularised lower incomplete gamma functions given by

$$\beta(x; a, b) = \frac{1}{\text{Be}(a, b)} \int_0^x t^{a-1} (1-t)^{b-1} dt, \quad 0 \leq x \leq 1,$$

$$\gamma(y; a) = \frac{1}{\Gamma(a)} \int_0^y t^{a-1} e^{-t} dt, \quad 0 \leq y < \infty,$$

with $a, b > 0$. We also denote by $\beta^{-1}(\cdot, a, b)$ and $\gamma^{-1}(\cdot, a)$ their corresponding inverses.

Our examples are motivated by transformations of the form $W := F^{-1}(V; \lambda)$, where F is the GP distribution function given by equation (1) and V is a random variable with support the unit interval $\mathbb{I} = [0, 1]$. The distribution function and density function of W are denoted by G and g , respectively. If $V \sim \text{uniform}(0, 1)$ distribution then $W \sim \text{GP}(\sigma, \xi)$ but if V is on \mathbb{I} with a distribution that contains the $\text{uniform}(0, 1)$ as a special case then more flexible distributions than the GP are produced for W .

2.2 Probability integral transform and new models

Let Ω_Y, Ω_V be two sample spaces. Define $Y : \Omega_Y \rightarrow \mathbb{R}$ to be a random variable with continuous probability density (distribution) function $f_Y(x; \eta)$ ($F_Y(x; \eta)$) parametrised over an m -dimensional vector of parameters $\eta \in \mathbf{H} \subseteq \mathbb{R}^m$. Let also $V : \Omega_V \rightarrow \mathbb{R}$ be a random variable with continuous probability density (distribution) function $f_V(v; \theta)$ ($F_V(v; \theta)$) parametrised over a d -dimensional vector of parameters $\theta \in \Theta \subseteq \mathbb{R}^d$, with $f_V(v; \theta) = 0$, if $v \notin [0, 1]$. We also assume the existence of a $\theta^* \in \Theta$ such that $f_V(v; \theta^*) = 1$, $\forall v \in [0, 1]$, i.e., a special case of V follows the $\text{uniform}(0, 1)$ distribution. Then, the distribution and density functions of the transformed random variable $F_Y^{-1}(V; \eta)$ are given by

$$K(x; \eta, \theta) = \Pr \{F_Y^{-1}(V; \eta) \leq x\} = F_V \{F_Y(x; \eta); \theta\}, \quad (3)$$

$$k(x; \eta, \theta) = K'(x; \theta, \eta) = f_V \{F_Y(x; \eta); \theta\} f_Y(x; \eta). \quad (4)$$

Therefore the distribution function $K(x; \eta, \theta)$ is a generalised distribution function for $F_Y(x; \eta)$ in the sense that $K(x; \eta, \theta^*) = F_Y(x; \eta)$, i.e., the distribution function $F_Y(x; \eta)$ is a special

case of $K(x; \boldsymbol{\eta}, \boldsymbol{\theta})$.

Equations (3) and 4 provide the basis of all subsequent generalizations we propose. The distribution function of V can be constructed in several different ways, one of which is the composition of a distribution function $L_1(\cdot; \boldsymbol{\psi})$, $\boldsymbol{\psi} \in \Psi \subseteq \mathbb{R}^{d-m}$, $\dim(\boldsymbol{\psi}) = d - m$, $d > 2m$, with the inverse of a distribution function $L_0(\cdot; \boldsymbol{\eta})$, $\boldsymbol{\eta} \in \mathcal{H} \subseteq \mathbb{R}^m$, $\dim(\boldsymbol{\eta}) = m$, where L_0 and L_1 are defined on the same support and L_0 is a special case of L_1 ($d > 2m$ is required for $\boldsymbol{\eta}$ to have lower dimension than $\boldsymbol{\psi}$), that is

$$F_V(v; \boldsymbol{\theta}) = \begin{cases} 0 & v < 0, \\ L_1\{L_0^{-1}(v; \boldsymbol{\eta}); \boldsymbol{\psi}\} & 0 \leq v \leq 1, \\ 1 & v > 1, \end{cases} \quad (5)$$

where $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^d$ consists of elements taken from the combined vector $(\boldsymbol{\psi}, \boldsymbol{\eta})$. When $\boldsymbol{\psi}$ and $\boldsymbol{\eta}$ have elements in common, the combined vector is interpreted as the vector consisting of the unique elements of $(\boldsymbol{\psi}, \boldsymbol{\eta})$ that span Θ . Note here that $\boldsymbol{\theta}$ is not necessarily equal to the combined vector $(\boldsymbol{\psi}, \boldsymbol{\eta})$ since common elements of $\boldsymbol{\psi}$ and $\boldsymbol{\eta}$, if any, are allowed to cancel in composition (5). For instance, when L_1 and L_0 are the distribution functions of the $\text{gamma}(\kappa, \sigma)$ and $\text{exponential}(\sigma)$ random variables, $\kappa > 0$, $\sigma > 0$, i.e., $L_1(x; \boldsymbol{\psi}) = \gamma(x/\sigma, \kappa)$ and $L_0(x; \boldsymbol{\eta}) = 1 - \exp\{-x/\sigma\}$, for $x > 0$, then equation (5) yields the distribution function $F_V(v; \boldsymbol{\theta}) = \gamma\{-\log(1-v), \kappa\}$, $0 \leq v \leq 1$. Here $\{\boldsymbol{\psi}, \boldsymbol{\eta}, \boldsymbol{\theta}\} = \{(\kappa, \sigma), \sigma, \kappa\}$ and $\{d, m, \dim(\Theta)\} = \{3, 1, 1\}$. Moreover, this distribution function reduces to the $\text{uniform}(0, 1)$ distribution when $\kappa = 1$, i.e., $\boldsymbol{\theta}^* = 1$.

Below we present three new probability density functions that are generalisations of the GP density and can be obtained by transformations of the form $F_Y^{-1}(V; \boldsymbol{\eta})$, where $F_Y = F$, $\boldsymbol{\eta} = \boldsymbol{\lambda}$ and V is a random variable that satisfies equation (5). Owing to the fact that each model extends the GP distribution in a parametric fashion, we refer to the new models as the extended GP (EGP) models and denote their density function by $g(x; \boldsymbol{\lambda}, \boldsymbol{\theta})$, for $x > 0$.

Example 1 Let $F_V(v; \boldsymbol{\theta}) = \beta \{1 - (1-v)^{|\xi|}, \kappa, |\xi|^{-1}\}$, $\boldsymbol{\theta} = (\kappa, \xi) \in (0, \infty) \times (-\infty, \infty)$. Then the transformed random variable $F^{-1}(V; \boldsymbol{\lambda})$ has probability density function given by

$$g(x; \boldsymbol{\lambda}, \boldsymbol{\theta}) = \begin{cases} \frac{|\xi|/\sigma}{\text{Be}(\kappa, |\xi|^{-1})} \left\{1 - (1 + \xi x/\sigma)_+^{-|\xi|/\xi}\right\}^{\kappa-1} (1 + \xi x/\sigma)_+^{-1/\xi-1} & \xi \neq 0, \\ \frac{\sigma^{-1}}{\Gamma(\kappa)} x^{\kappa-1} e^{-x/\sigma} & \xi \rightarrow 0. \end{cases} \quad (6)$$

Example 2 Let $F_V(v; \boldsymbol{\theta}) = 1 - \gamma\{-\log(1-v), \kappa\}$, $\boldsymbol{\theta} = \kappa \in (0, \infty)$. Then the transformed random variable $F^{-1}(V; \boldsymbol{\lambda})$ has probability density function given by

$$g(x; \boldsymbol{\lambda}, \boldsymbol{\theta}) = \begin{cases} \frac{\sigma^{-1}}{\Gamma(\kappa)} \left\{\xi^{-1} \log(1 + \xi x/\sigma)_+\right\}^{\kappa-1} (1 + \xi x/\sigma)_+^{-1/\xi-1} & \xi \neq 0, \\ \frac{\sigma^{-1}}{\Gamma(\kappa)} x^{\kappa-1} e^{-x/\sigma} & \xi \rightarrow 0. \end{cases} \quad (7)$$

Example 3 Let $F_V(v) = v^\kappa$, $\theta = \kappa \in (0, \infty)$. Then the transformed random variable $F^{-1}(V; \lambda)$ has probability density function given by

$$g(x; \lambda, \theta) = \begin{cases} \frac{\kappa}{\sigma} \left\{ 1 - (1 + \xi x/\sigma)_+^{-1/\xi} \right\}^{\kappa-1} (1 + \xi x/\sigma)_+^{-1/\xi-1} & \xi \neq 0, \\ \frac{\kappa}{\sigma} \left(1 - e^{-x/\sigma} \right)^{\kappa-1} e^{-x/\sigma} & \xi \rightarrow 0. \end{cases} \quad (8)$$

We write $W \sim \text{EGP1}(\kappa, \sigma, \xi)$, $W \sim \text{EGP2}(\kappa, \sigma, \xi)$ and $W \sim \text{EGP3}(\kappa, \sigma, \xi)$ when the density of a random variable W is given by expression (6), (7) and (8) respectively. In all examples in addition to the GP parameters σ and ξ there is a shape parameter $\kappa > 0$ that adds more flexibility in the main body of the density and does not alter its tail behaviour, i.e., all distributions have tail index $\xi \in \mathbb{R}$.

All models reduce to the GP density when $\kappa = 1$. More specifically, the EGP1 model can be viewed as an extended Snedecor's F_{ν_1, ν_2} distribution (Abramowitz and Stegun, 1965) with parameters $\nu_1 > 0$ and $\nu_2 \in \mathbb{R}$, that allows for negative ν_2 giving finite upper bound for this distribution. When $\xi > 0$ and $\kappa = \sigma$, the density reduces to the F_{ν_1, ν_2} distribution with $\nu_1 = 2\kappa$ and $\nu_2 = 2/\xi$. Additionally for $\xi > 0$, the EGP1 model is a well used loss distribution in actuarial science known in that literature as the *generalised Pareto* distribution (Hogg and Klugman, 1984; Klugman et al., 2008). The EGP1 model is an extension of this loss distribution for the case $\xi < 0$. The EGP2 model can be viewed as a model that generalises the GP density in a similar way to the gamma generalising the exponential distribution. Specifically, the GP distribution is the distribution of the random variable $(e^{\xi Y} - 1)\sigma/\xi$, where Y follows the exponential(1) distribution (Hosking and Wallis, 1987). Analogously, the EGP2 model is the distribution of the random variable $(e^{\xi Z} - 1)\sigma/\xi$, where Z follows the gamma($\kappa, 1$) distribution. Finally, the EGP3 distribution function is simply obtained by raising the GP distribution function $F(x; \lambda)$ to a power $\kappa > 0$.

2.3 Construction of extreme value models

Expression (5) represents a class of distribution functions F_V . However, unlike the extended models of § 2.2, the transformation $F_Y^{-1}(V; \lambda)$ does not always ensure that the resulting random variable has a tail index ξ for all values of θ . One such example can be obtained by taking $F_V = L_1 \circ L_0^{-1}$, with L_1 and L_0 being the distribution functions of Weibull(κ, σ) and exponential(σ) random variables, $\kappa, \sigma > 0$, i.e., $L_1(x) = 1 - \exp\{-(x/\sigma)^\kappa\}$. In this case, the transformed variable $F^{-1}(V; \lambda)$, has survival function \bar{G} given by

$$\bar{G}(x; \kappa, \lambda) = \exp \left[- \left\{ \xi^{-1} \log(1 + \xi x/\sigma)_+ \right\}^\kappa \right], \quad x > 0$$

which is a slowly varying function at ∞ for $\kappa < 1$, $\xi > 0$ and is therefore considered to be a ‘super-heavy-tailed’ distribution under this combination of parameters which means that the parameter ξ is no longer the tail index of this distribution. Hence, we proceed by characterising in Theorem 1 the class of distribution functions F_V under the assumption that $F_Y^{-1}(V; \eta)$ has tail index $\xi \in \mathbb{R}$ for all values of θ .

Theorem 1 Let $d > 2m$ where $d, m \in \mathbb{N}$ and consider the parameter vectors $(\eta, \psi, \theta) \in (\mathbb{H} \times \Psi \times \Theta) \subseteq (\mathbb{R}^m \times \mathbb{R}^{d-m} \times \mathbb{R}^d)$ with $\dim(\eta) = m$ and $\dim(\psi) = d - m$. Let $F_Y(x; \eta)$ be a twice differentiable distribution function of a random variable Y admitting a density function $f_Y(x; \eta)$. Let also V be a random variable with twice differentiable distribution function $F_V(v; \theta)$

so that its density function satisfies $f_V(v; \boldsymbol{\theta}) = 0$ when $v \notin [0, 1]$. Then the transformed random variable $F_Y^{-1}(V; \boldsymbol{\eta})$ has tail index $\xi \in \mathbb{R}$, if and only if, the distribution function of V can be represented by

$$F_V(v; \boldsymbol{\theta}) = 1 - \exp \left\{ - \int_0^v \frac{dz}{\xi f_Y\{F_Y^{-1}(z); \boldsymbol{\eta}\} C(z; \boldsymbol{\eta}, \boldsymbol{\psi})} \right\}, \quad 0 \leq v \leq 1, \quad (9)$$

where $C(z; \boldsymbol{\eta}, \boldsymbol{\psi}) = \int [s(z; \boldsymbol{\eta}, \boldsymbol{\psi}) / f_Y\{F_Y^{-1}(z); \boldsymbol{\eta}\}] dz$ and s is a real-valued function with $\lim_{z \rightarrow 1} s(z; \boldsymbol{\eta}, \boldsymbol{\psi}) = 1$ and $\int_0^1 [\xi f_Y\{F_Y^{-1}(z); \boldsymbol{\eta}\} C(z; \boldsymbol{\eta}, \boldsymbol{\psi})]^{-1} dz = \infty$.

A proof is given in Appendix. Theorem 1 gives the characterization of the class of distribution functions F_V from which can be constructed a new class of models, $F_Y^{-1}(V; \boldsymbol{\eta})$, that have tail index $\xi \in \mathbb{R}$. Under the assumption of $F_Y = F$ and $\boldsymbol{\eta} = \boldsymbol{\lambda}$, i.e., the case in the examples of § 2.1, we obtain the following.

Corollary 1 *Let V be a random variable as in Theorem 1. Then $F^{-1}(V; \boldsymbol{\lambda})$ has a tail index $\xi \in \mathbb{R}$ if and only if the distribution function of V is given by*

$$F_V(v; \boldsymbol{\theta}) = 1 - \exp \left\{ - \int_0^v \left[\xi(1-z)^{1+\xi} \int \frac{s(z; \boldsymbol{\lambda}, \boldsymbol{\psi})}{(1-z)^{1+\xi}} dz \right]^{-1} dz \right\}, \quad 0 \leq v \leq 1, \quad (10)$$

where s is a real-valued function with $\lim_{z \rightarrow 1} s(z; \boldsymbol{\lambda}, \boldsymbol{\psi}) = 1$ and $\int_0^1 [\xi(1-z)^{1+\xi} \int \frac{s(z; \boldsymbol{\lambda}, \boldsymbol{\psi})}{(1-z)^{1+\xi}} dz]^{-1} dz = \infty$.

Any real-valued function s with the specific properties of Theorem 1 would give rise to a valid distribution function F_V . As an example, consider the real-valued function $s(z; \boldsymbol{\lambda}, \boldsymbol{\psi})$ where $\boldsymbol{\psi} = \kappa \in (0, \infty)$, given by

$$s(z; \boldsymbol{\lambda}, \boldsymbol{\psi}) = z^{-\kappa} \{z^\kappa + \kappa - 1 + z^{\kappa+1} \xi - z(\kappa + \xi)\} / \{\kappa \xi(z-1)\}, \quad 0 \leq z \leq 1.$$

Then, equation (10) yields

$$F_V(v; \boldsymbol{\theta}) = 1 - \exp \left\{ - \int_0^v \left[\xi(1-z)^{1+\xi} \left\{ \frac{(1-z^\kappa)}{\kappa \xi z^{\kappa-1} (1-z)^{1+\xi}} + c \right\} \right]^{-1} dz \right\}, \quad c \in \mathbb{R}. \quad (11)$$

When $c = 0$, the distribution function $F_V(v; \boldsymbol{\theta}) = v^\kappa$ of Example 3 in § 2.2 is obtained. When $c > 0$ and $\xi \geq 0$, equation (11) yields a valid distribution function $F_V(v; \boldsymbol{\theta})$. However, c has to be necessarily equal to 0 for $F_V(v; \boldsymbol{\theta})$ to be a valid distribution function when $\xi < 0$.

2.4 Penultimate approximations

We have so far presented three examples from a general class of models that extends the GP distribution by incorporating additional parameters while preserving the tail index $\xi \in \mathbb{R}$. To characterise the deviation of the tail behaviour of the extended models from the GP distribution we examine the penultimate approximation of the tail index proposed by Smith (1987), i.e., we examine the rate of convergence of the three extended models given in § 2.2 to the GP survival function in limit expression (2). Let W be a random variable with twice differentiable distribution function $G(x)$ and density function $g(x)$. Denote also the reciprocal hazard function of W by $h_W(x) = [1 - G(x)]/g(x)$. Smith (1987) shows that for each u and $x > 0$ there exists $y \in [u, u + xh(u)]$ such that

$$\frac{1 - G(u + xh(u))}{1 - G(u)} = \{1 + h'(y)x\}_+^{-1/h'(y)}. \quad (12)$$

By virtue of expression (2) the scaled excess random variable $\{(W - u)/h(u)\} | W > u$ converges in distribution to the GP distribution if $\lim_{u \rightarrow x^G} h'(u) = \xi \in \mathbb{R}$. This is one form of the von Mises condition which is a necessary and sufficient condition for the convergence of the scaled excess of any random variable, with twice differentiable distribution function, to the GP distribution. Defining $u_n = G^{-1}(1 - 1/n)$, the penultimate approximation to the tail index in equation (12) is given in terms of n by $h'(u_n)$, as $n \rightarrow \infty$. Moreover, the rate of convergence to the GP distribution is given by $O\{|h'(u_n) - \xi|\}$.

Define $A_{\kappa,\xi} = \frac{(\kappa-1)}{1+1/|\xi|} \{|\xi|/\text{Be}(\kappa, 1/|\xi|)\}^{-|\xi|}$ and let $D_{\kappa,\xi} = (\xi - |\xi|)A_{\kappa,\xi}$ and $E_{\kappa,\xi} = |\xi|A_{\kappa,\xi}$ for $\xi \neq 0$. Table 1 shows the leading order terms from the penultimate approximations of the tail index for the EGP models. For $\xi \in [-1, 1]$, the EGP3 distribution admits the fastest rate of convergence whereas for $\xi \in [-1, 1]^c$ the EGP1 distribution has the fastest rate of convergence among the extended models. Irrespective of the value of ξ the EGP2 distribution has the slowest rate of convergence. Explicitly, for $\xi \neq 0$ and $\xi = 0$, the rate of convergence for the EGP1, EGP2 and EGP3 distributions is of order $\{n^{-|\xi|I(\xi>0)}n^{-2|\xi|I(\xi<0)}, (\log n)^{-1}, n^{-1}\}$ and $\{(\log n)^{-2}, (\log n)^{-2}, n^{-1}\}$, respectively. Here $I(\xi \in A)$ denotes the indicator function which takes the value 1 when $\xi \in A$ and 0 otherwise for any set $A \in \mathbb{R}$.

Table 1: Leading terms of threshold u_n and penultimate approximations $h'(u_n)$ for Examples 1–3 of § 2.2.

Model	u_n	$h'(u_n)(\xi \neq 0)$	$h'(u_n)(\xi \rightarrow 0)$
EGP1	$(\sigma/\xi) \left[\left\{ \frac{n \xi }{\text{Be}(\kappa, 1/ \xi)} \right\}^\xi - 1 \right]$	$\xi + n^{- \xi }D_{\kappa,\xi} - n^{-2 \xi }E_{\kappa,\xi}$	$-(\log n)^{-2}(\kappa - 1)$
EGP2	$(\sigma/\xi) \{n^\xi \Gamma(\kappa)^{-\xi} - 1\}$	$\xi + (\log n)^{-1}(\kappa - 1)$	$-(\log n)^{-2}(\kappa - 1)$
EGP3	$(\sigma/\xi) \{(\kappa n)^\xi - 1\}$	$\xi + n^{-1}(\kappa - 1)(\xi - 1)/(2\kappa)$	$-n^{-1}(\kappa - 1)/\kappa$

2.5 Statistics using extended GP Models

We propose the use of the EGP models as alternatives to the GP distribution for the modelling of the excess random variable $X - u | X > u$. Specifically, given a random sample \mathbf{x} we model the exceedances $\mathbf{x}_{>u} = \{x_i : x_i > u\} =: (x_1, \dots, x_{n_u})$ with the $\text{EGP}(\kappa, \sigma, \xi)$ family of distributions. Maximum likelihood is used to estimate the parameters (κ, σ, ξ) , i.e., maximum likelihood estimates satisfy

$$(\hat{\kappa}, \hat{\sigma}, \hat{\xi}) = \underset{(\kappa, \sigma, \xi) \in S}{\operatorname{argmax}} \ell(\kappa, \sigma, \xi | \mathbf{x}_{>u})$$

where $S = (0, \infty) \times (0, \infty) \times (-\infty, \infty)$ and $\ell(\kappa, \sigma, \xi | \mathbf{x}_{>u})$ denotes the log-likelihood of the parameters given the observed sequence of excesses of length $n_u = \#\{x_i > u\}$, i.e., for $\xi \neq 0$

$$\begin{aligned} \ell^{\text{EGP1}}(\kappa, \sigma, \xi | \mathbf{x}_{>u}) &= n_u \log \left\{ \frac{|\xi|/\sigma}{\text{Be}(\kappa, |\xi|^{-1})} \right\} + (\kappa - 1) \sum_{i=1}^{n_u} \log \left[1 - \{1 + \xi(x_i - u)/\sigma\}_+^{-|\xi|/\xi} \right] - \\ &\quad - (1/\xi + 1) \sum_{i=1}^{n_u} \log \{1 + \xi(x_i - u)/\sigma\}_+, \end{aligned}$$

$$\begin{aligned}\ell^{\text{EGP}2}(\kappa, \sigma, \xi | \mathbf{x}_{>u}) &= n_u \log \left\{ \frac{\sigma^{-1}}{\Gamma(\kappa)} \right\} + (\kappa - 1) \sum_{i=1}^{n_u} \log [\xi^{-1} \log \{1 + \xi(x_i - u)/\sigma\}_+] - \\ &\quad - (1/\xi + 1) \sum_{i=1}^{n_u} \log \{1 + \xi(x_i - u)/\sigma\}_+, \end{aligned}$$

$$\begin{aligned}\ell^{\text{EGP}3}(\kappa, \sigma, \xi | \mathbf{x}_{>u}) &= n_u \log (\kappa/\sigma) + (\kappa - 1) \sum_{i=1}^{n_u} \log \left[1 - \{1 + \xi(x_i - u)/\sigma\}_+^{-1/\xi} \right] - \\ &\quad - (1/\xi + 1) \sum_{i=1}^{n_u} \log \{1 + \xi(x_i - u)/\sigma\}_+. \end{aligned}$$

Inference for extreme quantiles is made via the T -observation return level x_T which is defined by the level that is exceeded on average once every T observations. The T -observation return level is the solution of $\Pr(X > x_T) = 1/T$. Under the assumption that the exceedances above a threshold u are well modelled by the EGP family of distributions and such that $x_T > u$, the T -observation return level for $\xi \neq 0$ is given by

$$\begin{aligned}x_T^{\text{EGP}1} &= u + \frac{\sigma}{\xi} \left[\left\{ 1 - \beta^{-1} \left(1 - (T\zeta_u)^{-1}, \kappa, |\xi|^{-1} \right) \right\}^{-\xi/|\xi|} - 1 \right], \\ x_T^{\text{EGP}2} &= u + \frac{\sigma}{\xi} \left[\exp \left\{ \xi \gamma^{-1} \left(\kappa, 1 - (T\zeta_u)^{-1} \right) \right\} - 1 \right], \\ x_T^{\text{EGP}3} &= u + \frac{\sigma}{\xi} \left[\left\{ 1 - (1 - (T\zeta_u)^{-1/\kappa}) \right\}^{-\xi} - 1 \right], \end{aligned}$$

where $\zeta_u = \Pr(X > u)$. Return level estimates are obtained by substituting the parameter values by their maximum likelihood estimates whereas standard errors and confidence intervals are derived by the delta method or from the profile likelihoods of the parameters.

Aside from the model fitting of the exceedances with the EGP family of distributions, additional diagnostics for the GP distribution can be obtained. In particular, extra insight about the convergence in expression (2) can be sought from the EGP models by testing the statistical hypothesis

$$H_0 : (\kappa, \sigma, \xi) \in S_0 \quad \text{vs} \quad H_1 : (\kappa, \sigma, \xi) \in S_1, \quad (13)$$

where $S_0 = \{(1, \sigma, \xi) : \sigma \in \mathbb{R}_+, \xi \in \mathbb{R}\}$ and $S_1 = \{(\kappa, \sigma, \xi) : \kappa \in \mathbb{R}_+ \setminus \{1\}, \sigma \in \mathbb{R}_+, \xi \in \mathbb{R}\}$. Given the sample of excesses $\mathbf{x}_{>u}$, the generalised log-likelihood ratio test statistic reads

$$\Lambda_{n_u}(\mathbf{x}_{>u}) = 2 [\sup \{\ell(\kappa, \sigma, \xi | \mathbf{x}_{>u}) : (\kappa, \sigma, \xi) \in S\} - \sup \{\ell(\kappa, \sigma, \xi | \mathbf{x}_{>u}) : (\kappa, \sigma, \xi) \in S_0\}], \quad (14)$$

where $S = S_0 \cup S_1$. From asymptotic likelihood theory as $n_u \rightarrow \infty$, Λ_{n_u} converges in distribution to the chi-squared with 1 degree of freedom under H_0 . Therefore, tests of the statistical hypothesis (13) can be made on the basis of the asymptotic distribution of Λ_{n_u} . Moreover, the limit expression (2) suggests that if the GP distribution is a reasonable model for the observed exceedances above a threshold u' , then exceedances above a higher threshold $u'' \geq u'$ should also follow the GP distribution. This argument suggests plotting $\hat{\kappa}$ against u and selecting the threshold as the lowest possible value at which $\hat{\kappa}$ is not significantly different from 1 and the estimated modified scale and tail index are constant for all $u'' > u'$.

3 Simulation Study

We illustrate the impact of the extended models on the tail estimation using normal simulated data. All comparisons are based on the root mean square error (RMSE) performance of a range of estimated extreme quantiles using various sample sizes for the simulations. Specifically, for each distribution 10000 samples of size $n = 100, 1000, 10000$ were generated. The GP, EGP1 and EGP2 distributions were fitted to the exceedances of each sample above a range of N equally spaced thresholds u_1, \dots, u_N , with $u_1 = \Phi^{-1}(1/n)$ and $u_N = \Phi^{-1}(1 - 30/n)$. Results obtained from the EGP3 model are not shown as they are similar to the EGP1 and EGP2 models. This grid was chosen such that u_1 and u_N correspond approximately to the minimum possible threshold, i.e., all data are above u_1 , and u_N is the threshold above which 30 data points are observed on average, respectively. At each threshold we computed Monte Carlo estimates of the RMSE of the T -observation return level estimate. For each sample size n used in the simulation study, we chose two different values of T , given by $T_i/n = 1.5, 5$, for $i = 1, 2$, corresponding to short and long extrapolations.

Figure 2 shows the RMSE output of the simulation study for the normal simulated data. Results illustrate improvement in inference using the EGP models over the GP model for both return level estimates and each sample size as the minimum RMSE is attained for the two EGP models, with their performance being almost indistinguishable at this value. More precisely, this improvement is largest in the small sample case ($n = 100$) where the optimal choice of the threshold according to the lowest RMSE is $u_1 = \Phi^{-1}(1/\{100\})$. This illustrates the advantage of fitting the EGP models to the whole data in small sample size cases instead of the GP distribution. For the T_1 -observation return level in the $n = 100$ case, the EGP estimates yield higher bias and lower variance than the GP estimates whereas for any other combination of sample size and return level, the EGP estimates have lower bias and either slightly lower or higher variance at the threshold where the minimum RMSE occurs. From Table 2 we also have that as the sample size increases, the absolute difference of the corresponding optimal thresholds and RMSE of the EGP distributions and the GP distribution diminishes. This is an expected phenomenon which is justified by the validity of the asymptotics of extreme value theory as sample size increases.

Table 2: Optimal thresholds for the estimation of the return levels for each sample size.

	T_1			T_2		
$u \backslash n$	100	1000	10000	100	1000	10000
u_{EGP}	-2.32	0.05	1.48	-2.32	0.51	1.44
u_{GP}	-0.33	0.45	1.70	-0.23	0.68	1.44

Figure 3 shows the Monte Carlo estimates as well as the estimated uncertainty of the shape parameter κ from the EGP2 model plotted against the threshold for all sample sizes. Note also that the estimates obtained from the EGP1 model are close to the EGP2 estimates and therefore are not shown here. All graphs illustrate the same feature, i.e., $\hat{\kappa}$ stabilises around the value 1 as the threshold u increases. Additionally, the minimum thresholds at which the value 1 is inside the sampling distribution of κ are similar to the optimal thresholds of Table 2 for the GP model, denoted by u_{GP} . This feature demonstrates the usefulness of this plot as an additional diagnostic for the GP modelling framework. The 95% pointwise confidence intervals are largest for small and large threshold values. This feature is explained by the greater dependence of parameters ξ and κ at low threshold values (revealed by the profile likelihood plots of ξ and κ that are not shown here) and the few data points at high threshold values.

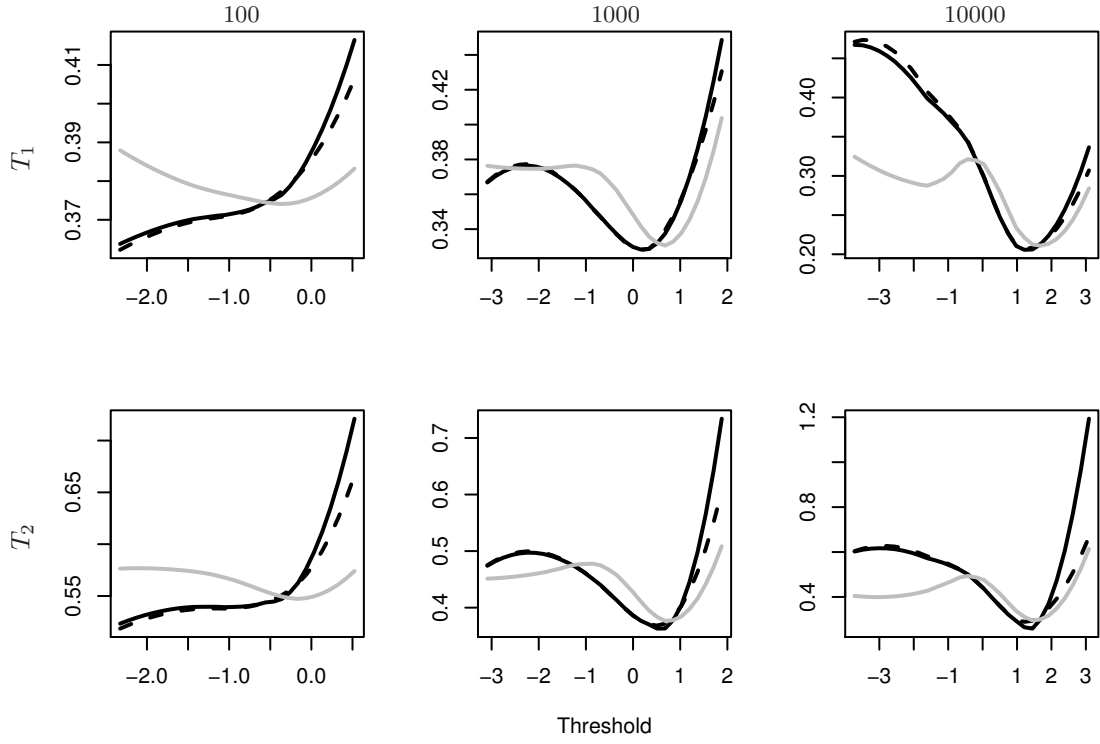


Figure 2: RMSE for the T_1 -(top row) and T_2 -observation (bottom row) return level estimates obtained from EGP1 (solid black), EGP2 (dashed black) and GP (grey). Columns correspond to the three different sample sizes $n = 100, 1000, 10000$ (from left to right).

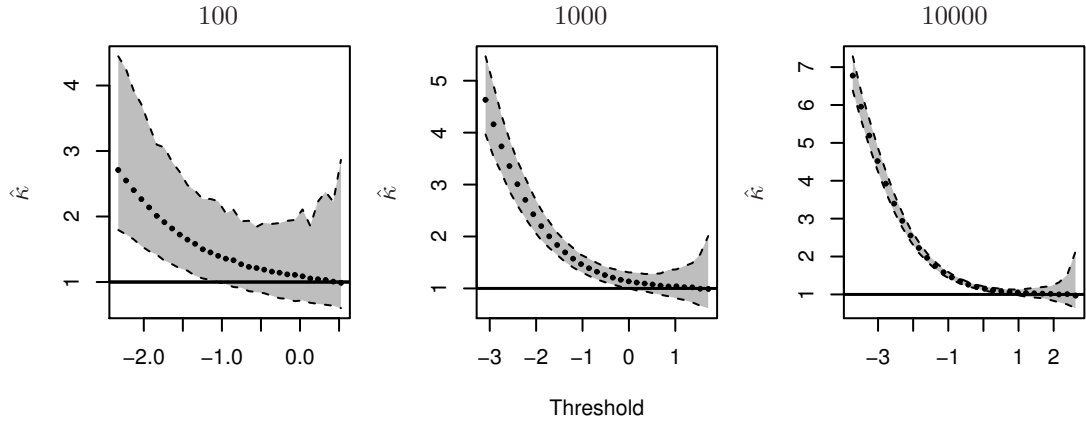


Figure 3: Monte Carlo estimates (median) of parameter κ (black dots) plotted against the threshold for each sample size. Grey-shaded areas correspond to the 95% pointwise equal tail confidence intervals.

4 Applications

4.1 River Nidd Data

We now analyse 154 exceedances of the threshold $65\text{m}^3\text{s}^{-1}$ by the River Nidd at Hunsingore Weir from 1934 to 1969 taken from NERC (1975). This data set constitutes the best known example with apparent difficulties in threshold selection and the modelling of the tail using the GP distribution, studied previously by Hosking and Wallis (1987), Davison and Smith (1990), Tancredi et al. (2006) and Wadsworth and Tawn (2011). Figure 4 shows the parameter stability plots from the EGP1 (left) and GP (right) models over a grid of thresholds $65.08, \dots, 88.61$ along with the histogram of the data. Threshold selection from the GP model based on the stability of the tail index and modified scale parameters is not straightforward. In contrast, the tail index and modified scale estimates from the EGP1 model appear to be stable over the plotted range of thresholds. Hence we select $u = 65\text{m}^3\text{s}^{-1}$ (all data points) for the fit of the EGP1 distribution. Moreover, the fact that $(\hat{\kappa}_{\text{EGP1}}, \hat{\sigma}_{\text{EGP1}}^*, \hat{\xi}_{\text{EGP1}})$ stabilise to values around

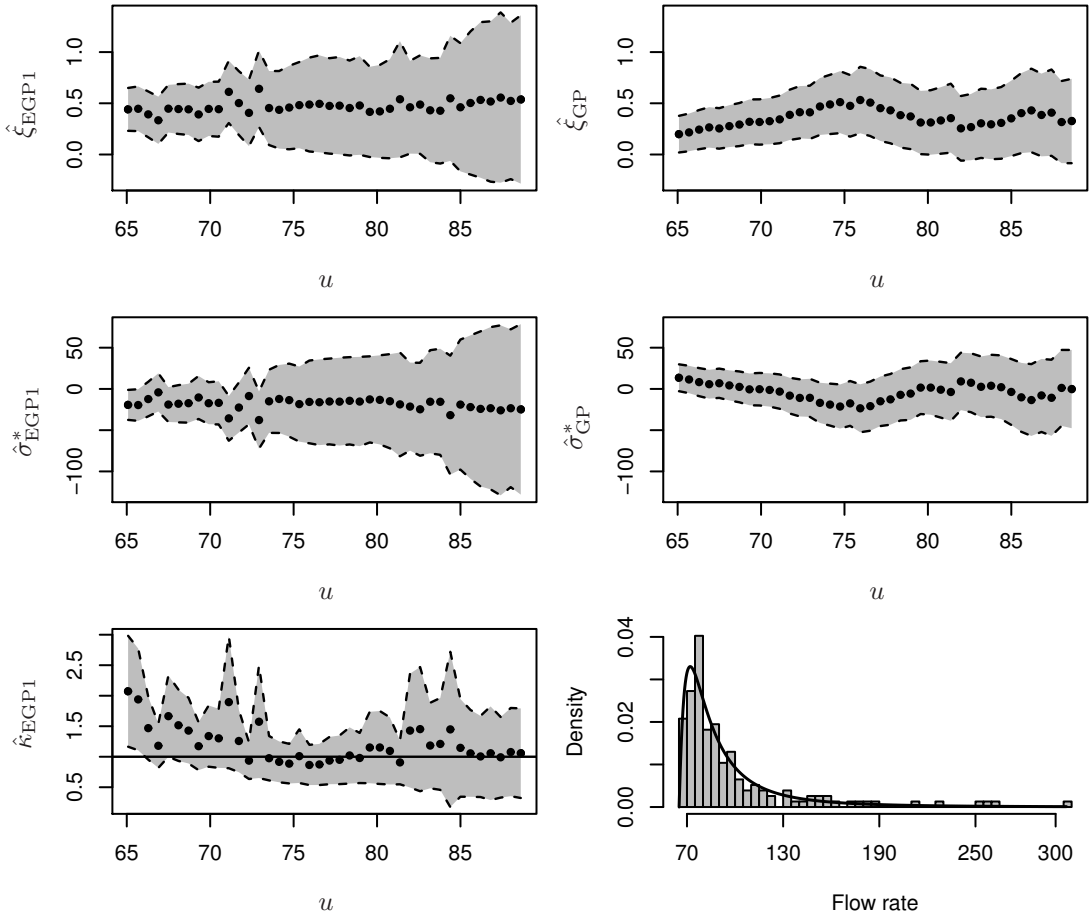


Figure 4: Maximum likelihood estimates and 95% pointwise equal-tail confidence intervals of tail index, modified scale and shape parameter (ξ, σ^*, κ) based on asymptotic normality: EGP1 (left column), GP (right column). Bottom right graph shows the histogram of the River Nidd data along with the estimated density (black solid line) from the EGP1 model fitted to the exceedances above $u = 65\text{m}^3\text{s}^{-1}$.

$(1, -16, 0.46)$ for the threshold values above 74 suggests that any threshold in this region is reasonable for the GP distribution. However, small deviations of $\hat{\kappa}_{\text{EGP1}}$ from the value 1 in this threshold region seem to have an impact on the stability of the GP estimates and the lowest

threshold where $\hat{\kappa}$ is very close to 1 is $75.3\text{m}^3\text{s}^{-1}$. This finding is also consistent with that of the Wadsworth and Tawn (2011) approach where they choose the value of $75\text{m}^3\text{s}^{-1}$. We thus select and $u = 75.3\text{m}^3\text{s}^{-1}$ for the GP distribution. Note also that the tail index and modified scale estimates from the EGP1 fitted above $u = 65\text{m}^3\text{s}^{-1}$ (0.44, -19) are similar to those obtained from the GP fitted above $u = 75.3\text{m}^3\text{s}^{-1}$ (0.48, -17).

To assess the impact on extrapolation, we look at the stability of return level estimates with respect to the choice of the threshold. Figure 5 shows return level estimates obtained from the EGP1 and GP models on the same grid of thresholds. Clearly, inference made on the basis of the EGP1 model yields much more stable results in comparison with the GP model. Return level estimates obtained from the EGP1 model gradually decrease with increasing threshold whereas estimates obtained from the GP model vary irregularly. This feature illustrates that the choice of threshold is less important for the Nidd data while using the EGP class of distributions.

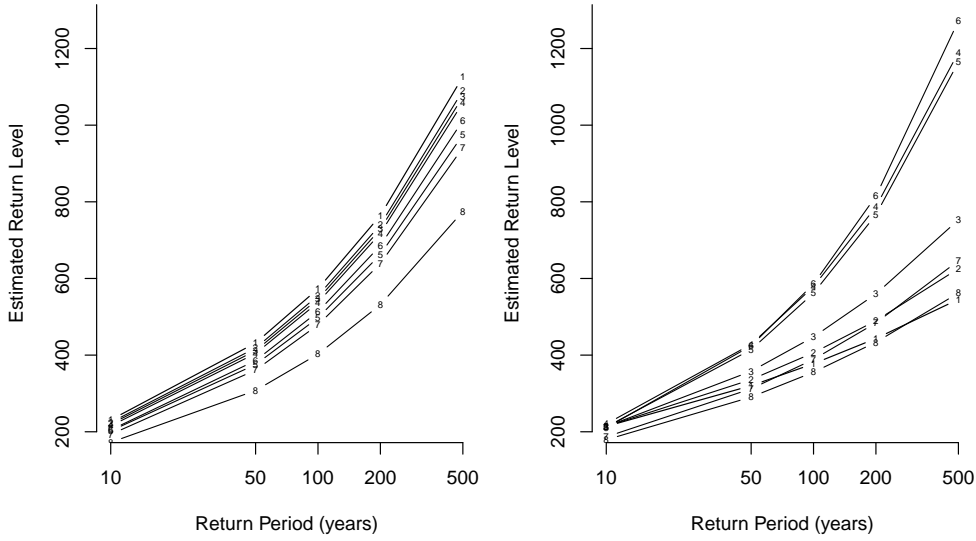


Figure 5: Estimates of 10-, 50-, 100-, 200- and 500-observation return level obtained from fitted EGP1 (left) and GP (right) models above the range of thresholds 65.08, ..., 82.6 coded here by numbers 1, ..., 8 respectively. Numbers 1 and 5 correspond to thresholds 65.08 and 75.3 respectively.

4.2 Pharmaceutical Application

We now return to the analysis of the residual bilirubin data shown in Figure 1. As already mentioned in § 1, the identification of liver toxic drugs is a multivariate extreme value problem in which the joint occurrence of extremes of residual bilirubin and other laboratory variables must be well modelled. However, as any multivariate extreme value analysis necessitates, the marginal extremes of these variables have to be modelled first. Southworth and Heffernan (2010) analysed the extremes of all laboratory variables taken from the same dataset with the GP modelling approach of Davison and Smith (1990), taking the threshold as the 70% quantile of the data. They found dose response relationships for all liver related laboratory variables other than residual bilirubin, justified by GP models with scale or tail index parameters linear in dose. Our primary objective in this analysis is to use the EGP1 distribution of § 2.2 to model the extremes of the residual bilirubin and to test for relationship with dose. Using the EGP models of § 2.2 allows the inclusion of more data points which might reveal evidence of rela-

tionship between residual bilirubin and dose, missed by Southworth and Heffernan (2010). To assess the relationship of residual bilirubin with dose we use generalised likelihood ratio tests between models that have dose dependent parameters and models with the same parameters across doses. The practice of pooling parameters and more specifically of the tail index in the extreme value modelling framework can be found in various applications including Coles and Tawn (1990); Cooley et al. (2007) and Davison et al. (2011) to name but a few.

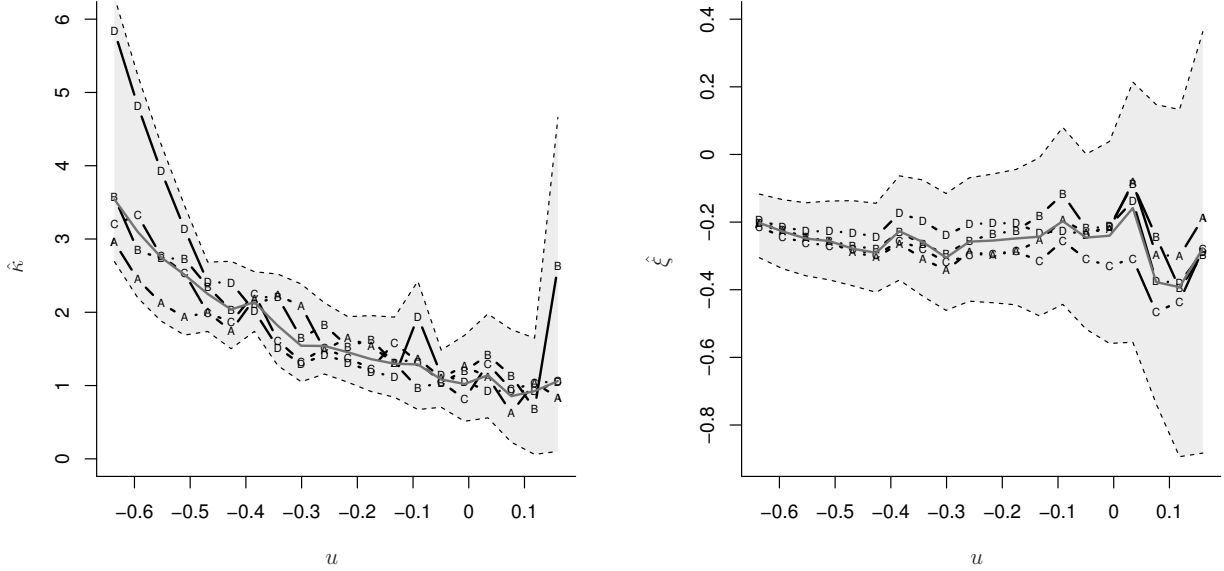


Figure 6: Left: estimated shape parameter for dose levels A, B, C, D (black lines) and under the assumption of common shape across doses (dark grey line). Right: estimated tail index for dose levels A, B, C, D under common shape across doses (black lines) and under the assumption of common shape and common tail index across doses (dark grey). Light grey areas correspond to a set containing all 95% pointwise confidence intervals based on asymptotic normality for the parameter estimates shown with black lines. The set is constructed by the minima (lower boundary) and maxima (upper boundary) of the confidence intervals.

Let $X^j - u | X^j > u$ be the excesses of the residual bilirubin variable over the threshold u at dose $j = A, B, C, D$. We initially fit the EGP1 model to the excesses over thresholds ranging from -0.65 (1%) to 0.15 (77%) by allowing separate shape, scale and tail index parameters for each dose, i.e., $X^j - u | X^j > u \sim \text{EGP1}(\kappa_j, \sigma_j, \xi_j)$, for dose j . The numbers in brackets are the corresponding sample quantiles of the combined data. The left plot of Figure 6 shows the maximum likelihood estimates $\hat{\kappa}_A, \dots, \hat{\kappa}_D$ over the threshold values. A feature revealed from this graph is that the estimated shape parameters appear to be similar across the doses for thresholds greater than -0.51. This is also supported by the generalised likelihood ratio test of the hypothesis $H_0 : (\kappa_A, \dots, \kappa_D) \in Q$ vs $H_1 : (\kappa_A, \dots, \kappa_D) \in Q^c$, where $Q = \{(\kappa_A, \dots, \kappa_D) \in \mathbb{R}_+^4 : \kappa_A = \dots = \kappa_D\}$ and Q^c is the complement of the set Q . Specifically, the generalised likelihood ratio test failed to reject the null hypothesis at all thresholds other than the threshold values below -0.51. Thus, we proceed to the analysis of the bilirubin data with the estimated common shape parameter shown with the dark grey line in the left plot of Figure 6. The right plot of Figure 6 shows the maximum likelihood estimates $\hat{\xi}_A, \dots, \hat{\xi}_D$ under the assumption of common shape across doses. In this case, the generalised likelihood ratio test failed to reject the null hypothesis of common tail index over dose at all thresholds. We found that the simplest model selected by generalised likelihood ratio tests is with common shape,

scale and tail index parameters for all doses. We also found similar results regardless of the order according to which the pooling of parameters was conducted. This suggests that there is no evidence of relationship between the residual bilirubin and dose for all thresholds greater than -0.51 , at the significance level of 5%. However, for thresholds below -0.51 there is evidence of a relationship with dose as indicated by the significant increase in the shape parameter estimate for dose D . This change indicates larger quantiles for dose D than for the other doses.

Figure 7 shows the quantile-quantile plots for the EGP1 and GP models with common shape, scale and tail index parameters among doses, fitted to the threshold exceedances above 0.10 (30%) and -0.13 (70%), respectively. The parameter estimates obtained from the EGP1 and GP fits are $(\hat{\kappa}_{\text{EGP1}}, \hat{\sigma}_{\text{EGP1}}, \hat{\xi}_{\text{EGP1}}) = (1.29, 0.25, -0.24)$ and $(\hat{\sigma}_{\text{GP}}, \hat{\xi}_{\text{GP}}) = (0.21, -0.27)$ respectively. Their corresponding standard errors are $(0.09, 0.02, 0.05)$ and $(0.01, 0.04)$. For the GP model we used Southworth and Heffernan (2010) choice of the 70% quantile which is consistent with the stability of the parameter estimates. For both models, the fit is good as the majority of the observed data points lie within the 95% pointwise tolerance intervals.

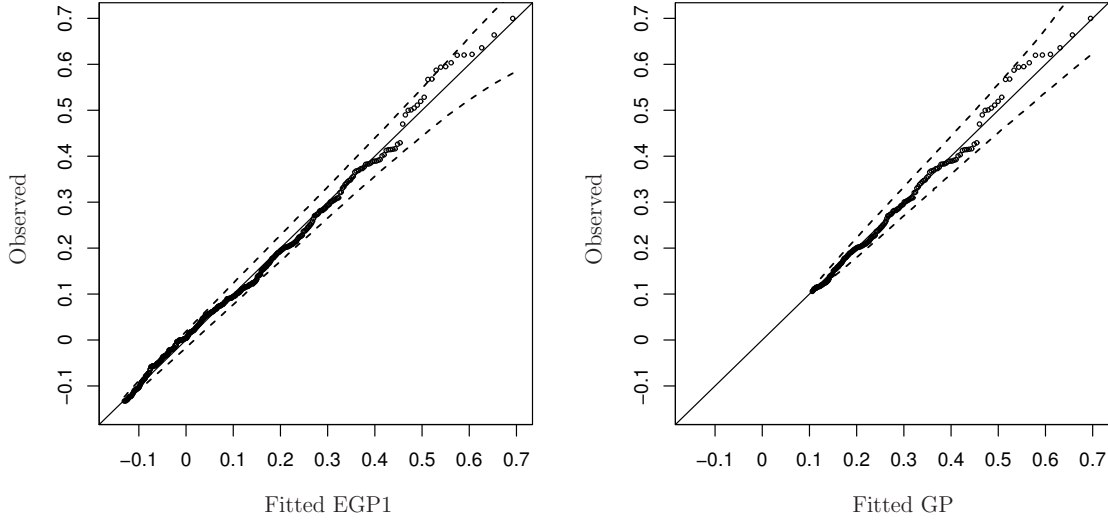


Figure 7: Quantile-quantile plots to assess the fit to the exceedances of the EGP1 (left) and GP (right) models. Dashed lines show the 95% pointwise tolerance intervals.

The best fitting EGP1 model has $\hat{\kappa}$ significantly different from 1, and hence provides evidence of a departure from the GP distribution at the selected threshold. However, above the respective thresholds used to fit the two models there is no apparent difference in the quality of the fits. The finding of no evidence of a dose effect in the EGP models is identical to findings of the previous GP analysis. Despite this failure to identify a dose effect for thresholds above -0.51 , we believe our analysis offers considerable benefits. Specifically, due to being able to substantially lower the threshold used relative to the GP analysis, larger sample sizes are used and thus the power of a test for dose effects in the residual bilirubin data is increased.

Acknowledgments

I. Papastathopoulos's work was carried under funding from Astrazeneca. We would particularly like to thank Harry Southworth of Astrazeneca for helpful discussions, suggestions and constructive comments on the analysis of the bilirubin data of § 4.2 and Ivar Struijker Boudier for carrying out and validating some of the numerical calculations of § 3.

A Proof of Theorem 1

Assume that F_Y can be represented by equation (9). Let $K(x; \boldsymbol{\eta}, \boldsymbol{\theta})$ and $k(x; \boldsymbol{\eta}, \boldsymbol{\theta})$ be the distribution function and density function of the transformed variable $W = F_Y^{-1}(V; \boldsymbol{\eta})$. Differentiability of f_Y and f_V implies that W will have tail index $\xi \in \mathbb{R}$ if the derivative of the reciprocal hazard function of W , $h'_W(x; \boldsymbol{\eta}, \boldsymbol{\theta}) = d/dx [\{1 - K(x; \boldsymbol{\eta}, \boldsymbol{\theta})\}/k(x; \boldsymbol{\eta}, \boldsymbol{\theta})]$, equals ξ as $x \rightarrow x^K = \sup \{x : K(x; \boldsymbol{\eta}, \boldsymbol{\theta}) < 1\}$ (Von Mises' condition). We have

$$\begin{aligned} h'_W(x; \boldsymbol{\eta}, \boldsymbol{\theta}) &= \xi C' \{F_Y(x); \boldsymbol{\eta}, \boldsymbol{\psi}\} \\ &= \xi s \{F_Y(x); \boldsymbol{\eta}, \boldsymbol{\psi}\} \\ &\rightarrow \xi \in \mathbb{R}, \quad \text{as } x \rightarrow x^K. \end{aligned}$$

To prove the converse, we assume that the random variable W has tail index ξ , i.e., $\lim_{x \rightarrow x^K} h'_W(x; \boldsymbol{\eta}, \boldsymbol{\theta}) = \xi$. In other words, there exists a real-valued function $s : \mathbb{R} \rightarrow \mathbb{R}$ with $\lim_{x \rightarrow x^{F_Y}} s \{F_Y(x); \boldsymbol{\eta}, \boldsymbol{\theta}\} = 1$ such that $h'_W(x; \boldsymbol{\eta}, \boldsymbol{\theta}) = \xi s \{F_Y(x); \boldsymbol{\eta}, \boldsymbol{\theta}\}$. Writing $h_W(x; \boldsymbol{\eta}, \boldsymbol{\theta}) = h_V \{F_Y(x; \boldsymbol{\eta}); \boldsymbol{\theta}\} / f_Y(x; \boldsymbol{\eta})$ we have

$$h'_V \{F_Y(x; \boldsymbol{\eta}); \boldsymbol{\theta}\} - \frac{f'_Y(x; \boldsymbol{\eta})}{f_Y^2(x; \boldsymbol{\eta})} h_V \{F_Y(x; \boldsymbol{\eta}); \boldsymbol{\theta}\} = \xi s \{F_Y(x); \boldsymbol{\eta}, \boldsymbol{\theta}\}.$$

The solution of this first order linear differential equation is given by

$$h_V \{F_Y(x; \boldsymbol{\eta}); \boldsymbol{\theta}\} = \xi f_Y(x; \boldsymbol{\eta}) \int s \{F_Y(x); \boldsymbol{\eta}, \boldsymbol{\theta}\} dx,$$

which is a separable differential equation with solution

$$F_V \{F_Y(x; \boldsymbol{\eta}); \boldsymbol{\theta}\} = 1 - \exp \left\{ - \int_0^{F_Y(x; \boldsymbol{\eta})} \frac{dF_Y(t; \boldsymbol{\eta})}{\xi f_Y(t; \boldsymbol{\eta}) \int s \{F_Y(t); \boldsymbol{\eta}, \boldsymbol{\theta}\} dt} \right\}.$$

Under the change of variable $z = F_Y(t; \boldsymbol{\eta})$, we have

$$F_V(v; \boldsymbol{\theta}) = 1 - \exp \left\{ - \int_0^v \frac{dz}{\xi f_Y \{F_Y^{-1}(z); \boldsymbol{\eta}\} \int \frac{s(z; \boldsymbol{\eta}, \boldsymbol{\theta})}{f_Y \{F_Y^{-1}(z); \boldsymbol{\eta}\}} dz} \right\}, \quad (15)$$

where $v = F_Y(x; \boldsymbol{\eta})$. By assumption $\boldsymbol{\theta}$ is an at most d -dimensional vector of parameters. Hence equation (15) implies the existence of a $(d-m)$ -dimensional vector of parameters $\boldsymbol{\psi} \in \Psi \subseteq \mathbb{R}^{d-m}$ such that expression (15) can be written as

$$F_V(v; \boldsymbol{\theta}) = 1 - \exp \left\{ - \int_0^v \frac{dz}{\xi f_Y \{F_Y^{-1}(z); \boldsymbol{\eta}\} \int \frac{s(z; \boldsymbol{\eta}, \boldsymbol{\psi})}{f_Y \{F_Y^{-1}(z); \boldsymbol{\eta}\}} dz} \right\},$$

and $(\boldsymbol{\psi}, \boldsymbol{\eta})$ span Θ .

References

- Abramowitz, M. and I. A. Stegun (1965). *Handbook of Mathematical Functions*. New York: Dover.
- Beirlant, J., G. Dierckx, Y. Goegebeur, and G. Matthys (1999). Tail index estimation and an exponential regression model. *Extremes* 2, 177–200.
- Beirlant, J., E. Joossens, and J. Segers (2009). Second-order refined peaks-over-threshold modelling for heavy tailed distributions. *J. Statist. Plann. Inference* 139, 2800–2815.
- Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*. London: Springer-Verlag.
- Coles, S. G. and J. A. Tawn (1990). Statistics of coastal flood prevention. *Phil. Trans. R. Soc. Lond. A* 332, 457–76.
- Cooley, D., D. Nychka, and P. Naveau (2007). Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Statist. Assoc.* 102, 824–840.
- Davison, A. C., S. Padoan, and M. Ribatet (2011). Statistical modelling of spatial extremes. *Statist. Science*. To appear.
- Davison, A. C. and R. L. Smith (1990). Models for exceedances over high thresholds. *J. R. Statist. Soc. B* 52, 393–442.
- FDA (2008). *Guidance for Industry-Drug Induced Liver Injury: Premarketing Clinical Evaluation*.
- Feuerverger, A. and P. Hall (1999). Estimating a tail exponent by modelling departure from a Pareto distribution. *Ann. Statist.* 27, 760–781.
- Frigessi, A., O. Haug, and H. Rue (2002). A dynamic mixture model for unsupervised tail estimation without threshold selection. *Extremes* 5, 219–235.
- Hogg, R. V. and S. A. Klugman (1984). *Loss distributions*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. New York: John Wiley & Sons Inc.
- Hosking, J. and J. Wallis (1987). Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics* 29, 339–349.
- Klugman, S. A., H. H. Panjer, and G. E. Willmot (2008). *Loss models: From data to decisions* (Third ed.). Wiley Series in Probability and Statistics. New York: John Wiley & Sons Inc.
- MacDonald, A., C. Scarrott, D. Lee, B. Darlow, M. Reale, and G. Russell (2011). A flexible extreme value mixture model. *Computational Statistics and Data Analysis* 55, 2137–2157.
- NERC (1975). *Flood Studies Report*. Natural Environment Research Council.
- Peng, L. (1998). Asymptotically unbiased estimators for the extreme-value index. *Statistics and Probability Letters* 38, 107–115.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *Ann. Statist.* 3, 119–131.

- Pickands, J. (1986). The continuous and differentiable domains of attraction in extreme value theory. *Ann. Probab.* 14, 996–1004.
- Smith, R. L. (1987). Approximations in extreme value theory. Technical Report 205, Department of Statistics, University of North Carolina, Chapel Hill.
- Southworth, H. and J. E. Heffernan (2010). *texmex: Threshold exceedences and multivariate extremes*. R package version 1.0.
- Tancredi, A., C. Anderson, and A. O’Hagan (2006). Accounting for threshold uncertainty in extreme value estimation. *Extremes* 9, 87–106.
- Wadsworth, J. and J. A. Tawn (2011). Likelihood-based procedures for threshold diagnostics and uncertainty in extreme value modelling. *J. R. Statist. Soc. B*. To appear.